

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus Botucatu

TATIANE CRISTINA RODRIGUES

**Impacto das mutações somáticas pontuais em regiões
reguladoras de Splicing em linhagens tumorais de mama.**

São Paulo - SP

2009

TATIANE CRISTINA RODRIGUES

**Impacto das mutações somáticas pontuais em regiões
reguladoras de Splicing em linhagens tumorais de mama.**

*Monografia apresentada junto ao curso de
Ciências Biológicas da UNESP - Universidade Estadual Paulista
“Julio de Mesquita Filho”, Campus de Botucatu,
na área de concentração de Genética, como requisito
parcial à obtenção do título de Bacharel.*

Orientador: Prof. Dr. Sandro José de Souza

Co-Orientadora: Dra. Júlia Pinheiro Chagas da Cunha

Co-Orientador: Dr. Pedro A. Favoretto Galante

São Paulo – SP

2009

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO
DA INFORMAÇÃO.
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: SELMA MARIA DE JESUS

Rodrigues, Tatiane Cristina.

Impacto das mutações somáticas pontuais em regiões reguladoras de Splicing em linhagens tumorais de mama / Tatiane Cristina Rodrigues. – Botucatu : [s.n.], 2010.

Trabalho de conclusão (bacharelado – Ciências Biológicas) – Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 2010

Orientador: Sandro José de Souza

Co-orientadores: Júlia Pinheiro Chagas da Cunha e Pedro A. Favoretto Galante

1. Câncer - Genética

Palavras-chave: Câncer; Mutação; Splicing

TATIANE CRISTINA RODRIGUES

**Impacto das mutações somáticas pontuais em regiões
reguladoras de Splicing em linhagens tumorais de mama.**

*Monografia apresentada junto ao curso de
Ciências Biológicas da UNESP - Universidade Estadual Paulista
“Julio de Mesquita Filho”, Campus de Botucatu,
na área de concentração de Genética, como requisito
parcial à obtenção do título de Bacharel.*

Orientador: Prof. Dr. Sandro José de Souza
Co-Orientadora: Dra. Júlia Pinheiro Chagas da Cunha
Co-Orientador: Dr. Pedro Favoretto Galante

Os componentes da banca de avaliação,
abaixo listados, consideram este trabalho aprovado.
COMISSÃO EXAMINADORA:

Prof. Dr. Sandro José de Souza
Instituto Ludwig de Pesquisa sobre o Câncer

Dra. Julia Pinheiro Chagas da Cunha
Instituto Ludwig de Pesquisa sobre o Câncer

Prof. Dr. Ney Lemke
Universidade Estadual Paulista “Júlio de Mesquita Filho” - campus Botucatu

APROVADO EM 10 DE DEZEMBRO DE 2009.

*Dedico este trabalho às minhas famílias:
a que me concebeu,
e a escolhida por mim para fazer parte dos meus dias.*

AGRADECIMENTOS

*“Se eu vi mais longe,
foi por estar de pé
sobre ombros de gigantes.”
(Isaac Newton)*

Agradeço à vida e à energia divina que nos rodeia, por absolutamente todas as oportunidades que me foram disponibilizadas, pelas possibilidades de erros, acertos e aprendizados, e, principalmente, pelas lindas surpresas presentes em cada dia.

Agradeço aos meus pais, irmão e avó, pelo amor, pelas incontáveis broncas, pelos sinceros carinhos, e pelo apoio às inúmeras situações aprontadas pela filha bióloga.

Agradeço ao Dr. Sandro, pela oportunidade que me foi dada, pelo voto de confiança, e pela orientação e incentivo. Agradeço à Dra. Júlia, por todos os ensinamentos (científicos e de vida), pelo enorme exemplo de dedicação, pela paciência infinita e pela amizade, essenciais ao desenvolvimento deste trabalho. Ao Dr. Pedro agradeço, por acreditar no meu potencial e me receber, pelo exemplo de postura profissional e pessoal, pela colaboração e pela amizade. Agradeço, por fim, a todos colegas do Laboratório e do Instituto Ludwig pela boa vontade em me receber, pelo apoio, pelos ensinamentos e pelo companheirismo.

Agradeço ao Dr. Ney Lemke pelas valiosas sugestões, orientação, interesse e revisão do presente trabalho.

Agradeço ao meu filho canino, por me demonstrar a existência e o valor do amor, companheirismo e da amizade incondicionais.

Agradeço aos meus loucos e lindos amigos, que sempre souberam estar por perto com tanto carinho, besteiras e força, pois, sem eles, com certeza, esta meta não teria sido cumprida, ou seria algo muito mais árduo e chato. Agradeço, em especial, a Alexandre (Produto), pela parceria incondicional e por me provar o enorme valor de uma amizade.

*“Tenho a impressão de ter sido uma criança
brincando à beira-mar, divertindo-me
em descobrir uma pedrinha mais lisa
ou uma concha mais bonita que as outras,
enquanto o imenso oceano da verdade
continua misterioso diante de meus olhos.”*
(Isaac Newton)

ÍNDICE

RESUMO.....	8
1. INTRODUÇÃO.....	9
2. OBJETIVOS.....	17
3. JUSTIFICATIVAS.....	18
4. MATERIAIS E MÉTODOS.....	19
5. RESULTADOS E DISCUSSÃO.....	24
6. CONCLUSÕES.....	41
APÊNDICE.....	43
BIBLIOGRAFIA.....	47

RESUMO

O evento de *splicing* alternativo tem como resultado a geração de diversos produtos a partir do precursor do RNA mensageiro de um único gene, sendo o responsável, assim, pelo aumento da variedade de transcritos e proteínas existentes em uma célula. Estima-se que cerca de 90% dos genes humanos estejam sujeitos a este tipo de processamento. O funcionamento adequado do processo de *splicing* depende do reconhecimento correto dos limites entre trechos intrônicos e exônicos pela maquinaria enzimática, que se dá através do reconhecimento de diversos sinais, como os sítios de *splicing* 3' e 5', o trato de polipirimidina, a seqüência “*branch*”, e pequenas seqüências presentes em exons e introns, próximas aos sítios de *splicing*, que promovem ou inibem a inclusão de trechos na fita de RNA madura. É fato comprovado por diversos estudos que mutações nas seqüências sinalizadoras de *splicing* podem modificar o padrão de processamento de um gene. Acredita-se que variações genéticas individuais possam modificar a susceptibilidade a diversas doenças, entre elas o câncer, que trata-se, atualmente, da doença que mais gera óbitos no mundo (13% do total). Recentemente, *Sjoblom et al.* (2006) e *Wood et al.* (2007) mapearam mutações não silenciosas encontradas em 1718 genes em linhagens de câncer de mama e colorretal. Neste trabalho, investigamos os efeitos dessas mutações somáticas presentes em câncer no padrão de *splicing* celular. Para tanto, nos focamos nas 201 mutações encontradas em quatro linhagens de câncer de mama (HCC1954, HCC1599, HCC1143 e HCC2157). A partir dos dados obtidos pela técnica de “*Exon Array*” (*Affymetrix*) e do mapeamento das mutações, foi realizada uma seleção dos genes aonde haviam mutações e eventos de *splicing* alternativos específicos a somente uma das linhagens celular, e cuja distância entre eles não ultrapasse 200 pares de bases, que é o trecho aonde há maior probabilidade de serem encontrados sinais de *splicing*. Sete genes (NLRC5, KIF16B, OXSM, KIAA1826, FEM1C, C18orf19 e EHBP1) foram submetidos a validação experimental, através de RT-PCR seguido de sequenciamento dos *amplicons*. Embora os dados obtidos através do “*Exon array*” sejam confiáveis, não conseguimos comprovar a existências de todos os transcritos alternativos. Foram encontrados quatro novos variantes de *splicing*, três deles presentes em todas as linhagens estudadas, e um supostamente específico a linhagem aonde a mutação se apresenta. Dentre os novos variantes encontrados, um deles já demonstra grande importância por apresentar um domínio enzimático trincado, o que deve comprometer a função da proteína sintetizada.

1. INTRODUÇÃO

1.1. Câncer e alterações genéticas envolvidas na carcinogênese

O câncer é, atualmente, a principal causa de morte no mundo: foram estimados 7,9 milhões de óbitos (o que representa 13% do número total) em 2007. A projeção é que o número de mortes devido ao câncer em todo o mundo continue aumentando, tendo-se estimado cerca de 12 milhões de vítimas fatais em 2030¹. Os principais tipos de câncer com maior mortalidade foram: pulmão (1,3 milhão), estômago (cerca de 1 milhão), fígado (662 mil), cólon (655 mil) e mama (502 mil). O câncer de mama é o segundo tipo de câncer mais freqüente no mundo e o mais comum entre as mulheres. A cada ano, cerca de 22% dos casos novos de câncer em mulheres são de mama. No Brasil, em 2008, houve um risco estimado de 51 casos deste tipo de câncer a cada 100 mil mulheres. Apesar de ser considerado um câncer de relativamente bom prognóstico se diagnosticado e tratado oportunamente, as taxas de mortalidade por câncer de mama continuam elevadas no Brasil, provavelmente porque a doença ainda é diagnosticada em estádios avançados. Na população mundial, a sobrevivência média após cinco anos é de 61%². Supõe-se que cerca de 30% dos óbitos devido ao câncer poderiam ser evitados³ e que, ao menos, um terço dos novos casos de câncer que ocorrem anualmente no mundo poderiam ser prevenidos².

A transformação de uma célula normal para um célula em desenvolvimento tumoral é um processo que envolve a aquisição de diversas alterações fenotípicas que se dão através do acúmulo quantitativo e qualitativo de mutações genéticas. Assim, o surgimento de instabilidades genômicas, mutações ou de certas variações durante o processo de expressão gênica têm o potencial de aumentar ou de refletir um início de desenvolvimento da carcinogênese. Conseqüentemente, tem-se a geração de fenótipos diferenciados e característicos a cada tipo de câncer⁴⁻⁹.

Acredita-se que variações genéticas individuais possam modificar a susceptibilidade a diversas doenças, assim como ao câncer. Desta maneira, a identificação e caracterização dos polimorfismos genéticos, como SNPs (“*Single-nucleotide polymorphisms*”) ^{10,11} e INDELS (“*insertions and deletions*”) merecem particular atenção ¹¹⁻¹³. Cada vez um número maior de polimorfismos genéticos está sendo caracterizado e seu papel na expressão gênica humana melhor estudado, através da caracterização de genomas individuais completos, como o de James D. Watson ¹⁴, Craig Venter ¹⁵, dentre outros estudos a respeito ¹⁶⁻¹⁸. Mais recentemente, o desenvolvimento de seqüenciadores de última geração aumentou a eficácia, diminuiu os custos e o tempo necessário para seqüenciar um genoma ou transcriptoma ¹⁹. As alterações mais freqüentes são os SNPs, que são definidos como mutações de base única que são encontradas em mais de 1% da população humana. Variantes de proteínas gerados por SNPs são responsáveis pela maioria da diversidade fenotípica humana, e existem casos relatados aonde um único SNP em um só gene é suficiente para ocasionar doenças ²⁰.

Diversas mutações somáticas presentes em tumores de mama, colorretal, pâncreas e glioblastoma foram identificadas através do seqüenciamento exônico de mais de 18 mil genes humanos pelos trabalhos de Sjöblom e Wood e colaboradores ^{7,8}, e a maioria dessas alterações eram desconhecidas. Algumas das principais observações desses trabalhos é que apesar de existirem mutações em diversos genes em cada tumor individual (aproximadamente 80 mutações em regiões codificadoras), menos de 15 devem ser primordiais para a iniciação e progressão deste, e, além disso, para cada tipo de tumor existem algumas vias metabólicas que apresentam seus genes mais freqüentemente mutados (por exemplo, em tumores de mama, diversos genes envolvidos na via PI3K estão freqüentemente mutados) ²¹.

Abaixo encontram-se destacadas as tabelas geradas pelos trabalhos de Sjöblom e Wood, aonde estão descritos o número, tipo e localização das mutações somáticas em linhagens tumorais por eles encontradas.

Tabela 1.a. Dados referentes a mutações somáticas em linhagens tumorais de mama e colorretal descritas e validadas pelo trabalho de Sjöblom e colaboradores.

Table 1. Summary of somatic mutations. Numbers in parentheses refer to percentage of total mutations.

Tumor	Discovery screen*			Validation screen†			Both screens combined		
	Colorectal	Breast	Total	Colorectal	Breast	Total	Colorectal	Breast	Total
Number of mutated genes	519	673	1149	105	137	236	519	673	1149
Number of mutations	574	733	1307	177	188	365	751	921	1672
Nonsynonymous mutations in coding sequences									
Missense	482 (84.0)	600 (81.9)	1082 (82.8)	126 (71.2)	145 (77.1)	271 (74.2)	608 (81.0)	745 (80.9)	1353 (80.9)
Nonsense	35 (6.1)	39 (5.3)	74 (5.7)	26 (14.7)	8 (4.3)	34 (9.3)	61 (8.1)	47 (5.1)	108 (6.5)
Insertion	3 (0.5)	3 (0.4)	6 (0.5)	2 (1.1)	2 (1.1)	4 (1.1)	5 (0.7)	5 (0.5)	10 (0.6)
Deletion	18 (3.1)	48 (6.5)	66 (5.0)	10 (5.6)	13 (6.9)	23 (6.3)	28 (3.7)	61 (6.6)	89 (5.3)
Duplication	17 (3.0)	2 (0.3)	19 (1.5)	3 (1.7)	12 (6.4)	15 (4.1)	20 (2.7)	14 (1.5)	34 (2.0)
Mutations in noncoding sequences									
Splice site‡	17 (3.0)	37 (5.0)	54 (4.1)	9 (5.1)	8 (4.3)	17 (4.7)	26 (3.5)	45 (4.9)	71 (4.2)
UTR§	2 (0.3)	4 (0.5)	6 (0.5)	1 (0.6)	0 (0.0)	1 (0.3)	3 (0.4)	4 (0.4)	7 (0.4)
Nucleotides successfully analyzed (Mb)	208.5	209.2	417.7	28.7	34.3	63.0	237.2	243.5	480.7
Mutation frequency (mutations/Mb)	2.8	3.5	3.1	6.2	5.5	5.8	3.2	3.8	3.5

*Coding and adjacent noncoding regions of 13,023 CCDS genes were sequenced in 11 colorectal and 11 breast cancers. †Genes mutated in the discovery screen were sequenced in 24 additional tumor samples of the affected tumor type. ‡Intronic mutations within 4 bp of exon/intron boundary. §Mutations in untranslated regions (UTR) within 4 bp 5' of initiation codon or 4 bp 3' of termination codon. ||Nucleotides with Phred quality score of at least 20.

(Fonte: Sjöblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. Science 314, 268-274 (2006).)

Tabela 1.b. Dados de análise das mutações somáticas em linhagens de câncer de mama e colorretal descritas por Wood e colaboradores.

Table 1. Summary of somatic mutations. UTR, untranslated region. ND, not determined because synonymous mutations were not evaluated in the RefSeq genes analyzed in (5).

Tumor type	Screen	Gene set	Mutated genes	Coding changes						Noncoding changes	Total mutations
				Missense	Nonsense	Insertion	Deletion	Duplication	Synonymous	Splice site or UTR	
Colorectal cancers	Discovery	This study	325	237	14	0	8	0	93	12	364
		All RefSeq	848	722	48	4	27	18	ND	30	942
	Validation	This study	88	81	9	1	2	2	30	6	131
		All RefSeq	183	197	34	4	14	5	ND	15	299
Breast cancers	Discovery	This study	460	304	26	2	28	1	131	14	506
		All RefSeq	1137	909	64	5	78	3	ND	53	1243
	Validation	This study	62	52	3	0	3	0	19	2	79
		All RefSeq	167	153	11	2	15	2	ND	7	209

(Fonte: Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. Science 318, 1108-1113 (2007).)

1.2. Controles da expressão gênica: *splicing* e *splicing alternativo*

As diversas atividades exercidas por cada célula, assim como a própria homeostase celular, são controladas pela expressão gênica. O termo expressão gênica refere-se ao processo em que a informação contida em um determinado gene, inicialmente dada como a seqüência de DNA, é processada em um produto gênico funcional, como proteínas ou os diversos tipos de RNA. A modificação em qualquer uma das etapas desse processo tem potencial de ocasionar uma expressão gênica diferencial^{22,23}. O DNA genômico não direciona a síntese proteica por si só, mas através do uso do RNA como molécula intermediária. Sob um aspecto generalizado, em todas as células, a expressão da informação genética é dada da seguinte forma: o DNA é molde para a síntese de RNA, processo este chamado de transcrição, e o RNA especifica a síntese dos polipeptídios que formarão as proteínas, que se trata do processo de tradução.

De acordo com suas necessidades, a célula efetua transcrição de diferentes trechos de DNA de uma de suas fitas. Os genes são unidades situadas em espaçamentos irregulares ao longo da seqüência de DNA, que, a fim de conceber a exatidão de posição necessária à transcrição, geralmente contêm sinais específicos: sítios promotores e de terminação. Tais sinais são reconhecidos pelas enzimas responsáveis por parte do processo, as RNA polimerases. Os processos de iniciação, alongamento e término da transcrição envolvem uma complicada série de transições estruturais entre moléculas de proteínas, DNA e RNA. Ao contrário do que ocorre em grande parte dos procariotos, nos eucariotos todo o reconhecimento de sinais e a produção de uma fita de RNA pela RNA polimerase são somente o primeiro passo da transcrição. Nestes organismos, o precursor do RNA mensageiro, primeiro produto da transcrição, está sujeito a uma série de eventos de processamento antes que sua saída do núcleo seja permitida e haja sua tradução em proteínas²⁴.

Esse processamento consiste basicamente em 3 etapas:

1. a adição, pela RNA polimerase, de cerca de 100 até 250 resíduos de adenilato (AMP) à extremidade 3' da seqüência gerando a chamada “cauda poli A”, que tem como funções facilitar o deslocamento do RNA pela célula, estabilizar o mRNA e servir de sítio identificador da molécula, facilitando a tradução;
2. a ligação de um nucleotídeo especializado (trifosfato 7-metilguanosina) formando uma espécie de proteção “cap” à extremidade 5', que protege o transcrito do ataque de exonuclease, facilita o deslocamento do RNA para o citoplasma e tem papel conhecido na ligação da subunidade 40S dos ribossomos no mRNA;
3. o *splicing*, um dos passos fundamentais para o processamento do pré-mRNA, que consiste na remoção de introns e na junção dos exons remanescentes através de reações de transesterificação, catalisadas por um grande complexo enzimático ribonucleoproteico chamado Spliceossomo. A catálise ocorre associada a uma série de rearranjos e combinações entre o spliceossomo, a fita de mRNA e os cofatores do processo²⁵.

A dependência de inúmeros fatores para que a ocorrência da catálise permite, simultaneamente, uma maior regulação de cada etapa do processo e uma flexibilidade para a adaptação à enorme variedade de introns encontrados numa típica célula eucariótica²⁶. Qualquer alteração ao longo de todo este evento do pré-mRNA pode alterar quantitativa e qualitativamente o produto polipeptídico resultante da tradução do RNA^{24,27}.

Em suma, o evento de *splicing* pode ser descrito como uma série de etapas, dependentes de sinais específicos, que efetua parte do processamento do precursor do RNA mensageiro, visando guiar e regular a produção de proteínas que suprirão as diversas necessidades celulares ou teciduais existentes em um dado momento. Como variações no mecanismo de *splicing* têm potencial para gerar diferentes transcritos a partir de um único gene, este processo é o maior contribuinte tanto para a diversidade proteômica como para o controle dos expressão gênica, que são fatores regulados diferencialmente de acordo com o

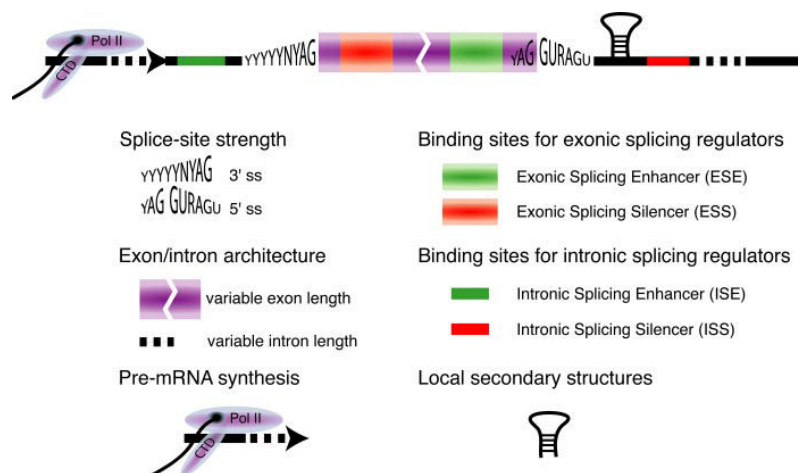
tipo do tecido, função, estágio de desenvolvimento e, inclusive, em casos de acometimento de doenças²⁸⁻³⁰. Dada sua devida importância, o *splicing* necessita, assim, de um controle refinado que assegure que o evento ocorra primorosamente.

A maquinaria enzimática que catalisa a reação de *splicing* do pré-mRNA, comumente denominada Spliceossomo, talvez seja a mais complexa maquinaria celular^{25,31}. Ela é constituída de 5 pequenas moléculas de RNA chamadas de “*small nuclear RNAs*” ou snRNAs (U1, U2, U4, U5 e U6) e cerca de 300 proteínas distintas que ainda não têm função totalmente caracterizada no evento. A catálise ocorre associada a uma série de importantes rearranjos entre as moléculas de RNA e de proteínas do Spliceossomo²⁵.

O funcionamento adequado do processo de *splicing* depende do reconhecimento correto dos limites entre introns e exons pela maquinaria enzimática. A inclusão ou exclusão de um dado trecho se dá por intermédio de um recrutamento de fatores que funcionam como adaptadores entre o pré-mRNA e o Spliceossomo. Os detalhes do processo que permite esta identificação precisa pela maquinaria de *splicing* ainda não são bem compreendidos, entretanto é comprovada a existência de diversos sinais que auxiliam na identificação dos limites entre regiões exônicas e intrônicas. Alguns de tais sinais têm procedência nos nucleotídeos adjacentes ao sítio de *splicing*, e estão longe de serem randômicos, consistindo em duas seqüências consenso de fácil distinção com aproximadamente 9 bases no sítio 5' e 15 bases no sítio 3'³². Dentro destas seqüências consenso, em praticamente todos os exons, existem dinucleotídeos intrônicos GT e AG localizados respectivamente nos sítios 5' e 3' de *splicing*. Há também fortes indícios da existência de sinais provenientes de seqüências internas aos introns, como o “*branch point*”, que quase sempre contém uma adenosina bem conservada entre as espécies e está situado entre 20 a 40 nucleotídeos *upstream* ao sítio de *splicing* localizado na extremidade 3' da molécula^{25,33}.

Entretanto, somente essas seqüências não são suficientes para definir o limite exato entre um exon e um intron, sendo necessária a presença concomitante de seqüências adicionais internas aos segmentos que auxiliam na ativação ou inibição da inclusão de um determinado trecho no transcrito de mRNA. Esses elementos adicionais podem ser divididos em quatro categorias, de acordo com sua função: “*exonic splicing enhancers*” (ESEs), “*exonic splicing silencers*” (ESSs), “*intronic splicing enhancers*” (ISEs) e “*intronic splicing silencers*” (ISSs)³⁴. Tais elementos reguladores de *splicing* servem como sítios de ligação de certas famílias protéicas, como a família SR – rica em serina e guanina – para os ESSs, que recrutam subunidades do Spliceossomo, incentivando ou inibindo a inclusão de trechos de pré-mRNA no processo.

Figura 1.a. Esquema dos elementos reguladores e demais sinais de *splicing*

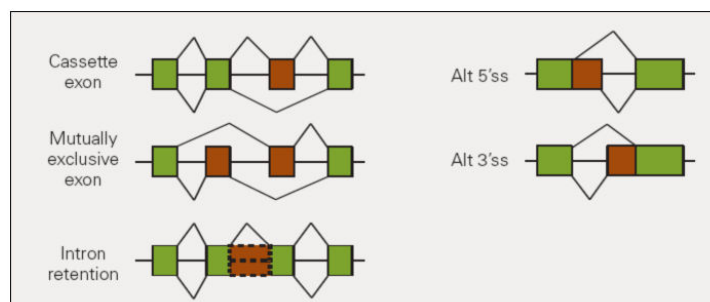


(Fonte: Hertel, K. J. Combinatorial control of exon recognition. *J.Biol.Chem.* 283, 1211-1215 (2008).).

A inclusão ou exclusão de seqüências em um trecho de mRNA de forma variável a partir de eventos de *splicing* é denominada *splicing* alternativo²⁶. Tradicionalmente, considera-se a existência de 5 tipos de *splicing* alternativo: “*exon skipping*” ou “*cassette exon*”, “*mutually exclusive exons*”, “*intron retention*”, “*alternative 5' splice site*” e “*alternative 3' splice site*”³⁵ (vide Figura 2). Evidências encontradas por intermédio de

análise de transcritos levam a conclusão de que os diferentes tipos de *splicing* alternativo podem, por vezes, ocorrer também de maneira concomitante^{36,37}.

Figura 2. Tipos de *splicing* alternativo.



(Fonte: *Identifying and Validating Alternative Splicing Events. Affymetrix. Technical Note. 2006.*)

Estima-se que cerca de 90% dos genes humanos estejam sujeitos ao processo de *splicing* alternativo³⁸. Em mais de 80% dos casos, o produto gerado é uma proteína modificada³⁸ (funcional ou não) alterando a variedade do proteoma^{39,40}, mas seu efeito pode ainda ser neutro. Assim, este é um importante processo molecular que modula a expressão gênica eucariótica, e é fundamental para o aumento da diversidade proteômica^{30,36,41}.

O fim do projeto de seqüenciamento do genoma humano, ao revelar a existência de um número menor de genes que o esperado para a espécie, renovou o interesse no fenômeno de *splicing* alternativo⁴². Em estimativa, o *splicing* alternativo afeta a expressão de cerca de 92 a 94% dos genes humanos⁴¹. Um ponto de grande importância a ser investigado nesta questão seria a determinação do conjunto de *motifs* de seqüências genéticas que incitam a ocorrência de eventos de *splicing*, que propiciará, então, a predição dos padrões de *splicing* de qualquer transcrito primário a partir de sua seqüência.

2. OBJETIVOS

2.1. Objetivo geral:

- Investigar se mutações de um par de bases, específicas a determinadas linhagens celulares tumorais e localizadas em regiões próximas às bordas de *splicing*, têm a capacidade de influenciar a ocorrência de eventos de *splicing* alternativo, gerando transcritos diferentes dos gerados pelo gene quando a mutação está ausente.

2.2. Objetivos específicos:

- Explorar o acometimento de alterações genômicas, mais especificamente mutações de um par de bases que se encontram próximas a sítios de *splicing* alternativo, encontradas especificamente em uma das seguintes linhagens celulares: HCC1954, HCC1143, HCC1599 e HCC2157;
- Mapear mutações genômicas de um par de bases próximas a eventos de *splicing* alternativo previstos por *exon array* por intermédio de ferramentas de bioinformática;
- Selecionar genes candidatos promissores com a proposta para validação;
- Validar a existência e expressão de transcritos alternativos para cada candidato selecionado por intermédio de RT-PCR, usando oligonucleotídeos específicos que amplifiquem o trecho do transcrito supostamente envolvido em evento de *splicing* alternativo;
- Estudar o impacto das mutações genômicas pontuais supra citadas no padrão de *splicing* dessas células.

3. JUSTIFICATIVAS

Diversos estudos foram realizados analisando os efeitos indutores de variações genéticas, especialmente SNPs, em eventos de *splicing*, mostrando que alterações em um único nucleotídeo podem alterar o padrão de processamento de um gene^{43,44}. Outros estudos comprovam também a existência de diversas doenças causadas, agravadas, ou que trazem como consequência um evento de *splicing* alternativo^{45,46}.

O acometimento por câncer envolve a ocorrência de uma série de alterações genéticas específicas ou não à doença, como mutações. Dependendo do tipo e do local de sua ocorrência, alterações gênicas podem abolir, modificar ou criar sítios ou seqüências reguladoras de *splicing*. A conexão entre câncer, alterações genéticas específicas da doença e mudanças nos níveis de expressão de transcritos variantes é algo bem fundamentado^{29,33,37,47-50}. Porém, atualmente, não há estudos que investiguem se há uma relação entre a existência de mutações pontuais em trechos que possivelmente atuam como seqüências reguladoras de *splicing* e a existência de eventos de *splicing* alternativo, supostamente em decorrência dessas mutações.

Assim, sendo, o objetivo do presente estudo é investigar se mutações de um par de bases, específicas a determinadas linhagens celulares tumorais e localizadas em regiões próximas às bordas de *splicing*, têm a capacidade de influenciar a ocorrência de eventos de *splicing* alternativo, gerando transcritos diferentes dos gerados pelo gene quando a mutação está ausente. Os resultados deste estudo tendem a dar indícios da influência das mutações no processamento gênico, além de permitir a caracterização de variantes de *splicing* presentes, especificamente ou não, em câncer de mama.

4. MATERIAIS E MÉTODOS

4.1. Análise de dados por intermédio de ferramentas de bioinformática

A seqüência do genoma humano (NCBI build 36.1) está disponível em bancos de nosso laboratório e foi obtida do site do UCSC Genome Bioinformatics (<http://genome.ucsc.edu>). No total, temos 26.281 seqüências do tipo *Reference Sequence (release 22)*; <http://www.ncbi.nlm.nih.gov/RefSeq/>). Foram criados então 4 bancos de dados com as seguintes informações:

Banco 1: Descrição das mutações descritas por Sjöblom e Wood: suas posições genômicas e quais são as alterações de bases;

Banco 2: Dados de cDNA humano: nome e localização cromossômica de genes, exons e introns a partir do genoma referência;

Banco 3: Dados provenientes do *Exon array* (Affymetrix®): ID probe set, posição genômica da probe;

Banco 4: Dados gerados pela análise do Exon array: intensidade da probe, normalização dos dados e índice comparativo “Splicing Index”⁵¹.

As informações presentes nos bancos de dados foram cruzadas, permitindo a identificação dos casos aonde a mutação está próxima a um região contendo possíveis transcritos variantes identificados pelo *Exon Array*. Através desta análise dos dados foram selecionadas as regiões gênicas candidatas a validação.

4.2. Amostras de linhagens de células tumorais

Amostras das linhagens de células tumorais HCC1599, HCC2157, HCC1143 e HCC1954 foram obtidas de American Type Culture Collection (ATCC).

4.3. Cultura e propagação de linhagens celulares

A propagação das células em cultura foi feita utilizando-se dos meios de crescimento recomendados pelo fabricante. Para as linhagens HCC1954, HCC1599 e HCC1143 foi utilizado o meio RPMI (GIBCO®) com adição de 10% Soro Fetal Bovino (Cultilab®) e 1% de antibiótico gamicina. Para a linhagem HCC2157, o meio foi o ACL-4 (constituído da mistura de 1:1 do meio “Ham’s F12” (Sigma Aldrich®) com adição de L-glutamina a 2,5 mM, com o meio “Dulbecco’s Modified Eagle’s” (Sigma-Aldrich®) ajustado a conter bicarbonato de sódio a concentração de 1,2 g/L; e suplementos adicionais: insulina 0,02 mg/mL, transferrina 0,01 mg/mL, selenite de sódio 25 nM, hidrocortisona 50 nM, Fator de Crescimento Epidermal 1 ng/mL, etanolamina 0,01 mM, fosforiletanolamina 0,01 mM, triiodotironina 100 pM, albumina bovina 0,5% (peso/volume), HEPES 10 mM, piruvato de sódio 0,5 mM, L-glutamina 2 mM). O meio ACL-4 foi ainda suplementado com Soro Fetal Bovino 10%. As condições de subcultura foram adotadas de acordo com as características de crescimento de cada linhagem, atentando para recomendações do fabricante e procedimentos padrão utilizados na sala de

cultura do Instituto Ludwig. De uma maneira geral, o meio das células foi trocado a cada 2-3 dias e as células aderidas foram tripsinizadas quando atingiam a confluência de 90%.

4.4. Extração, eletroforese e quantificação de RNA total

O RNA total foi isolado usando TRIzol® (Invitrogen™), de acordo com o protocolo sugerido pelo fabricante. A integridade dos RNAs foi visualizada por eletroforese, aplicando-se 1µg de RNA total em gel de agarose 1%. O gel de agarose foi preparado para 25 mL a cada placa de eletroforese, e teve a composição de 1% agarose em 1x TAE (Tris-Acetato-EDTA). Antes de ser aplicado no gel, o RNA foi desnaturado a 65°C durante 5 minutos, sendo mantido em condição desnaturante em tampão de amostra contendo uréia (uréia 7 M, 30% glicerol e corante azul). Todos os procedimentos foram realizados com materiais próprios para uso exclusivo em RNA. A coloração do gel foi feita utilizando GelRed™ Nucleic Acid Gel Stain (Biotium®) e o mesmo foi visualizado através de incidência de luz ultravioleta. Foram considerados íntegros os RNAs que apresentaram as bandas correspondentes aos RNAs ribossômicos 28S e 18S evidentes, sendo a primeira duas vezes mais intensa que a segunda. A quantificação do RNA foi feita pela leitura em espectrofotômetro a um comprimento de onda equivalente a 260 nm, considerando-se que 1 DO 260 nm equivale a 40 µg/ml de RNA. A fim de estabelecer um parâmetro de estimativas de contaminação do RNA foram feitas leituras também a comprimentos de onda de 280 nm e 230 nm. A relação entre as leituras a 260 e 280 nm estimaram o grau de contaminação do RNA por proteínas e a relação entre as leituras a 260 e 230 nm estimaram a contaminação por sais e outros componentes orgânicos.

4.5. Síntese de cDNA

A reação da transcriptase reversa foi realizada utilizando-se o kit “First-Strand Synthesis using SuperScript™ II RT” (Invitrogen®). A reação consistiu em 2 µg de RNA, 1µL de oligonucleotídeo dT (12-18) 500 µg/mL, 1µL de 10mM dNTP mix e água DEPC até completar 12 µL de reação. As reações foram então mantidas a 65°C por 5 minutos e, em seguida, centrifugadas rapidamente. Então, foram adicionados 4 µL de First Strand Buffer 5x, 2 µL de DTT 0,1 M e 2 unidades de inibidor “RNase OUT” por reação. A seguir as reações foram incubadas a 42°C por 2 minutos, e adicionado 1 µL da enzima SuperScript II (Invitrogen). A síntese de cDNA foi realizada por 2 horas a 42°C seguida da inativação da polimerase por 15 minutos a 70°C. O cDNA foi estocado a 4°C.

4.6. Reações em cadeia da polimerase (PCR).

4.6.1. Reações de PCR para gene controle Notch2

Os iniciadores (*primers*) utilizados estão identificados na Tabela 4.a desta seção. Os produtos amplificados foram visualizados a partir de uma eletroforese em gel de 8% poliacrilamida corado com nitrato de prata. Protocolos detalhando os delineamentos das reações de PCR controle estão listados na seção Apêndice.

4.6.2 Reações de PCR para os genes de interesse

O cDNA sintetizado foi utilizado como molde em reações de PCR com iniciadores específicos para cada gene candidato selecionado. Os oligonucleotídeos sintéticos iniciadores (primers) foram desenhados usando as seqüências apropriadas para cada caso (vide Tabela 4.b desta seção), com o auxílio do software Oligotech® e foram obtidos através da IDT – Integrated DNA Technologies. Os produtos amplificados nas reações foram visualizados a partir de uma eletroforese em gel de 8% acrilamida corado com nitrato de prata. Cada um dos protocolos de PCR está disponibilizado na seção Apêndice.

4.7. Gel de poliacrilamida

Para cada placa 7 mL de gel com foi preparada uma solução de poliacrilamida 8% (29:1, acrilamida: bisacrilamida) em TBE (Tris Borato EDTA – Tris 0,1 M e EDTA 2 mM), 35 µL de APS 20% (Persulfato de amônio) e 7 µL de TEMED (Tetrametilenodiamino). As amostras foram aplicadas em um tampão Ficcol 6x O tampão de corrida foi o TBE 1x. Os géis foram corados usando o nitrato de prata, para tal foi adicionado uma solução fixadora (2,25 L de água pura, 2,5 mL de etanol, 19,5 mL de ácido acético) e 2 mL de nitrato de prata 10% ao gel, que permaneceu por 10 minutos. O gel foi lavado 2 vezes com água destilada e, em seguida, o gel foi imerso numa solução NaOH 3% com 1 mL de formaldeído 37%, lavado com água destilada novamente, e colocada novamente a solução fixadora, quando, por fim, o gel é retirado para secagem.

4.8. Clonagem das regiões genômicas

Foi utilizado o kit “InsTAclone PCR cloning” (Fermentas®) para clonagem de fragmentos de interesse resultantes de PCR, utilizando o vetor pT257 R/T. A ligação do inserto ao vetor foi realizada seguindo a proporção de 3:1 (inserto : vetor), utilizando 1,5 unidades da enzima T4 DNA ligase e tampão 5x do kit. A reação de ligação foi realizada por 12 horas, seguida de uma diálise por 20 minutos. A transformação por eletroporação foi realizada utilizando 3 µL de ligação para cada 50 µL de bactéria competente (E. coli DH10B). As bactérias foram plaqueadas em placas de LB-ágar contendo antibiótico ampicilina (0,1mg/mL), em estufa a 37°C.

4.9. Reações de MiniPrep

As reações de mini-preparação para isolamento de plasmídeo bacteriano carregador de gene clonado foram realizadas utilizando o kit “Wizard® Plus SV Minipreps - DNA Purification System” (Promega®). Uma única colônia de bactérias foi inoculada em 5 mL de meio LB com ampicilina (0,1 mg/mL) e a solução foi incubada a 37°C por 12 horas com agitação. Então, 5 mL da cultura bacteriana foram centrifugados a 10.000 g por 5 minutos, com posterior descarte do sobrenadante e do excesso de

meio. A fim de ressuspender o pellet, foram adicionados 250 μL de “Cell Resuspension Solution” (composto de 50 mM de Tris-HCl, 10 mM de EDTA e 100 $\mu\text{g}/\text{mL}$ de RNase A) e a solução foi incubada por 5 minutos. A seguir, foram adicionados 10 μL de “Alkaline Protease Solution” e a solução foi novamente incubada por até 5 minutos. Foram, então, adicionados 350 μL de “Neutralization Solution” (composto de hidróclorido de guanidina 4,09 M, acetato de potássio 0,759 M e ácido acético glacial 2,12 M) e imediatamente misturado por inversões do tubo. A solução foi centrifugada por 10 minutos em velocidade máxima, e o sobrenadante foi transferido para uma coluna de separação por decantação, aonde foi centrifugada por mais um minuto em velocidade máxima. A coluna foi lavada duas vezes com 750 μL e 250 μL de “Column Wash Solution” (acetato de potássio 162,8 mM, Tris-HCl 22,6 mM e EDTA 0,109 mM). O DNA plasmidial que ficou retido na coluna foi, então, eluído com 100 μL de água livre de nucleases. A solução com DNA plasmidial foi armazenada a -20°C .

4.10. Digestão dos vetores

A solução de DNA plasmidial resultante do Miniprep foi clivada com as enzimas de restrição específicas para *E. coli*: XbaI e SalI (Promega), segundo instruções do fabricante. A reação de digestão consistiu em uma solução contendo 16,3 μL de água ultra pura, 2 μL de tampão 10x específico para cada enzima de restrição, 0,2 μL de BSA acetilado, 1 μg de DNA plasmidial e 0,5 μL de enzima de restrição (5 unidades). Foi realizada incubação a 37°C por 4 horas. Então, 4 μL de 6x “Loading buffer” foram adicionados e o produto de digestão foi analisado em gel de agarose.

4.11. Extração de material genético de gel de agarose

Para extração e purificação de DNA, com tamanhos variando entre 70 pares de bases e 1 kilobases, a partir do gel de agarose em TBE (Tris-Borato-EDTA), foi utilizado o kit “QIAquick Gel Extraction Kit Protocol” (Qiagen®), de acordo com as informações sugeridas pelo fabricante. Foram adicionados 3 volumes de tampão “QG” a cada 1 volume de gel, incubado a 50°C até a dissolução do gel, e então foi adicionado 1 volume de isopropanol. A solução foi colocada na coluna de separação proveniente do kit e centrifugada, descartando-se o sobrenadante. Foi então adicionado mais tampão “QG” sobre o filtro da coluna de separação, seguido de outra centrifugação. Para a lavagem, foi adicionado 0,75 mL de tampão “PE” à coluna de separação, seguido de outra centrifugação, quando então todo o sobrenadante foi descartado. Foi efetuada uma última centrifugação a 17.900 g por 1 minuto, e posteriormente a coluna foi transferida para um tubo de 1,5 mL limpo. Para eluir o DNA foi adicionado 50 μL de tampão “EB” (10mM Tris-HCl, pH 8,5) ao centro da membrana de separação e centrifugado por 1 minuto.

4.12. Seqüenciamento

O seqüenciamento de bases dos ácidos nucléicos foi realizado utilizando o kit de marcador fluorescente “Big Dye” (Applied Biosystems®) seguindo recomendações do fabricante. Foram utilizados 200 ng de cada produto de amplificação. Foram realizadas duas reações separadamente para cada produto de PCR e em cada uma foi utilizado 0,5 µL de iniciador (forward ou reverse), 1 µL de tampão 10X e o volume final foi completado para 10µL com água ultra pura. Foi realizada uma ciclagem com as seguintes temperaturas: 95°C por 1 minuto e 30 segundos, 35 ciclos de 95°C por 18 segundos, 58°C por 12 segundos e 60°C por 4 minutos, e finalizada a 16°C. O produto da reação de marcação fluorescente foi precipitado utilizando-se uma solução de 80 µL de isopropanol durante 15 minutos, centrifugada a 30.000g, e o pellet foi lavado com uma solução de 150 µL de etanol 70%. A seguir foi adicionado 10 µL de formamida ultra pura e a solução foi agitada vigorosamente. A desnaturação foi obtida após 5 minutos a 95°C, seguida 2 minutos no gelo. As reações foram aplicadas no aparelho seqüenciador ABI3100 (Applied Biosystems®) segundo instruções do fabricante.

Tabela 4.a. Iniciadores para PCR dos genes de controle

Gene	Iniciador Forward	Iniciador Reverse
Notch 2	5' CCAACCAGTTCTCCTGC 3'	5' GTGGTTAGGGCAGTCATC 3'

Tabela 4.b. Iniciadores para PCR dos genes selecionados

Gene	Iniciador Forward	Iniciador Reverse
NLRC5	5' CGGAAGCAGTGCAAGAAG 3'	5' AAGAACTCCTGCAGGCTG 3'
KIF16B	5' AGGCAACAGCGTGAAGAAC 3'	5' TTCTGCTGAACCAGGTCC 3'
OXSM	5' CCGACTGTGGAGAAGTGTC 3'	5' CTCTTGGCACATAAGCAGC 3'
KIAA1826	5' AGATTCAGGGATGCAGGTG 3'	5' TTCCAAGGACGGTTTCTCTG 3'
FEM1C	5' GTGTTTCGATGGCCATTTG 3'	5' CCTTGGATGCAGCTTAAG 3'
C18ORF19	5' TGCTCTGCTGAGTGACAC 3'	5' TTCCTTTGCAACACACTGG 3'
EHBP1	5' GCCAGAAGCATTGTGACC 3'	5' CTCCTCGGTAGTTCTTTG 3'

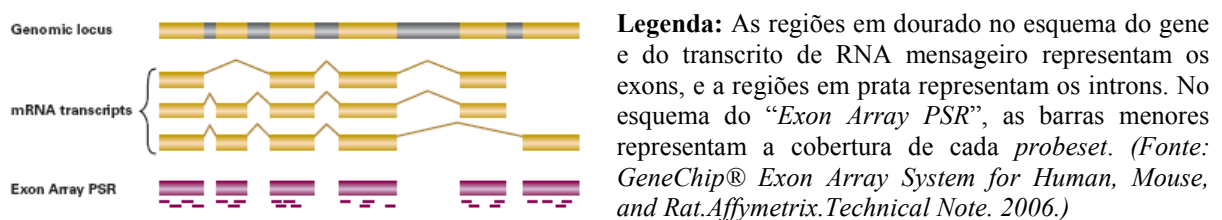
5. RESULTADOS E DISCUSSÃO

5.1. Ensaio de “*Exon Array*” e seleção dos genes candidatos

Com o objetivo de avaliar o impacto de mutações de um par de bases em eventos de *splicing* alternativo em câncer, analisamos o padrão de *splicing* de 4 linhagens tumorais através de ensaios de “*Exon Array*” (Affymetrix®). Tais linhagens tumorais (HCC1954, HCC1143, HCC2157 E HCC1599 (ATCC®)) foram escolhidas por apresentarem diversas mutações documentadas através do seqüenciamento de 18 mil genes, e por apresentarem características tumorais semelhantes segundo classificação TNM (“*Classification of Malignant Tumours*” - União Internacional contra o Câncer), como tipo, grau e localização do tumor.

A ocorrência de *splicing* alternativo diferencial entre as linhagens pode ser determinada pela diferença quantitativa de expressão de certo trecho genômico. Para a predição das regiões aonde estariam acontecendo eventos de *splicing* alternativo diferenciais para uma ou mais linhagens em relação as demais, utilizamos os dados provenientes da técnica de “*Exon Array*” (Affymetrix®). Nesta técnica, os desenhos das *probes* são feitos de acordo com vários bancos de dados de bioinformática como *Esembl*, *NCBI*, *RefSeq*. As *probes* são construídas em todos os exons mapeados de todos os genes conhecidos. O *RefSeq* possibilita a escolha de seqüências consideradas mais confiáveis e menos confiáveis do genoma para as quais os *chips* são desenhados possibilitando a cobertura de todos os exons conhecidos. Com cerca de 1,4 milhões de *probesets*, aproximadamente 4 *probes* por exon, tais *chips* possibilitam distinguir diversas isoformas de um gene.

Figura 5.a. Esquema das regiões de cobertura do “*Exon Array*”.



Os dados obtidos através da técnica de “*Exon Array*” foram normalizados e comparados de acordo com o sugerido pelo fabricante, como demonstram as figuras a seguir.

Figura 5.b. Cálculo de normalização da intensidade da *probeset*. Cálculo do índice comparativo “*Splicing Index*”.

Gene-level normalized intensity (NI).		Splicing Index Value (SI).		
$\text{Gene-level Normalized (NI) Intensity} = \frac{\text{Probe set intensity}}{\text{Expression level of the "gene"}}$		NI	Sample 1	Sample 2
		$\frac{\text{Probe set intensity}}{\text{Gene level}} = \frac{500}{500} = 1.0$	$\frac{600}{6000} = 0.1$	
		Sample 1 has 10x higher inclusion level		
		SI	Sample 1 NI	Sample 2 NI
		$\text{Splicing Index} = \log_2 \frac{1.0}{0.1} = +3.32$		

(Fonte: *Identifying and Validating Alternative Splicing Events. Affymetrix. Technical Note. 2006.*)

Cada *probeset* do “*Exon Array*” é específica a um pequeno trecho gênico. A comparação entre as intensidades de *probesets* entre diferentes linhagens, quando anteriormente normalizados com a expressão do gene em questão da linhagem (vide *figura 5.b*), possibilita uma predição segura de que há trechos sendo mais ou menos inclusos na fita de RNA mensageiro maduro em determinada amostra. Para facilitar as comparações desses níveis de inclusão entre diferentes linhagens foi utilizado o índice “*Splicing Index*” (vide *figura 5.b*). O valor obtido no “*Splicing Index*” foi, então, utilizado para comparar o padrão de *splicing* de um certo trecho genômico entre diferentes amostras, e prever em quais locais podem estar ocorrendo inclusões ou exclusões de trechos genômicos no RNA mensageiro maduro. Selecionando os casos aonde o valor do índice “*Splicing Index*” de um *probeset* fosse maior que 1,5 ou menor que -1,5, ou seja, quando a expressão relativa de um trecho genômico era 3 vezes maior em uma linhagem que na outra comparada, obtivemos um esboço dos locais genômicos aonde haveria possibilidade de estar ocorrendo um evento de *splicing* alternativo.

Para prever quais mutações estariam diretamente afetando o *splicing*, cruzamos a seleção de dados obtidos a partir da técnica de “*Exon Array*” com as mutações selecionadas.

Selecionamos, então, mutações somáticas de um par de bases, encontradas por *Sjöblom*⁸ e *Wood*⁷ que se localizavam a, no máximo, 200 pares de bases das bordas de *splicing* dos exons dos genes, pelo fato de estas serem as que apresentam uma maior chance de afetarem o funcionamento do mecanismo de *splicing*. A visualização de um possível evento de *splicing* alternativo específico restrito à linhagem que contém a mutação e localizado em proximidade a esta alteração genética caracterizaram algumas regiões genômicas como potenciais alvos para nosso estudo.

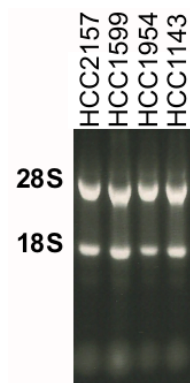
Partimos, então, para uma análise manual de cada um dos candidatos que preencheram tais requisitos, a fim de filtrar os casos que pudessem estar relacionados a um falso positivo. Nesta fase de análise, observamos se o local de ocorrência dos eventos poderia ser real. Uma das maneiras que utilizamos para isso foi verificar se o valor de “*Splicing Index*” entre as linhagens tendia a zero, e, somente quando a linhagem que apresentava a mutação estava sendo relacionada, era gerado um índice maior que $|1,5|$. A seguir, observamos todos os transcritos descritos nos bancos de dados públicos, para verificar a existência de transcritos com maior ou menor inclusão de um dado trecho genômico, o que aumentaria a chance de que a alteração do índice observada seja real. Outra observação realizada foi se as *probesets* com índices que apresentavam tais valores estavam localizadas suficientemente próximas (menos de 200 pares de bases) das mutações relatadas na literatura.

Selecionamos, então, 7 genes que continham regiões com fortes indícios de apresentarem uma mutação somática de um par de bases afetando um evento de *splicing* próximo a ela, para validar por RT_PCR a presença de eventos de *splicing* alternativo. Os genes selecionados foram: NLRC5, KIF16B, OXSM, KIAA1826, FEM1C, C18orf19 e EHBP1.

5.2. Detecção dos transcritos constitutivos e variantes

Realizamos a cultura, propagação celular, e extraímos o RNA total celular de cada uma das linhagens tumorais. A integridade do RNA das linhagens foi verificada através de eletroforese, aonde podem ser claramente visualizadas as bandas ribossômicas 18S e 28S, demonstrando que a extração não degradou as fitas de RNA (vide *figura 5.d*).

Figura 5.d. Eletroforese do RNA total celular.



Legenda: Acima: Linhagens HCC2157, HCC1599, HCC1954 e HCC1143.
À esquerda: bandas ribossômicas 18S e 28S.

Baseado nas seqüências referência disponibilizadas por “*UCSC Genome Bioinformatics*” (<http://genome.ucsc.edu>), desenhamos pares de iniciadores (*primers*) específicos para cada gene, na tentativa de flanquear o evento a ser visualizado. Assim, os produtos amplificados gerados inicialmente por *splicing* alternativo teriam tamanhos diferentes dos *amplicons* do transcrito constitutivo e seriam facilmente distinguidos em uma eletroforese visualizada em gel.

Para um controle interno da eficiência da síntese de cDNA, este foi amplificado com iniciadores desenhados especificamente para uma porção terminal (3') do gene Notch 2 (Notch homolog 2). Como o Notch2 é um gene relativamente extenso, se sua porção terminal for amplificada é um bom indicador de que a síntese de cDNA ocorreu conforme o desejado, transcrevendo as fitas de mRNA por inteiro. A amplificação de um fragmento de aproximadamente 300 pares de bases indicou a eficiência da síntese de cDNA.

O cDNA sintetizado foi utilizado como molde em reações de PCR com iniciadores específicos para cada gene candidato selecionado. Diversas condições da reação de PCR foram testadas para cada caso a fim de escolhermos a que proporcionasse maior eficiência com menor amplificação de produtos indesejados. Como pode ser visto na figura 5.e, variamos as concentrações de Magnésio, as temperaturas de anelamento dos *primers* e a quantidade de cDNA utilizadas.

Figura 5.e. Condições de PCR testadas para cada gene.

NLRC5		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C		⊗		
58°C		XX⊗	X		
60°C					

FEMIC		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C			X	X⊗	⊗
60°C		X			X

KIF16B		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C			XX	X⊗	⊗
60°C		X	X		

C18orf19		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C			X	XX⊗	XX
60°C		X	X		X

OXSM		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C		XX⊗	X⊗		
60°C		XX	X		

EHBPI		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C			XX⊗		
60°C		X	XX	X	X

KIAA1826		Concentração MgCl ₂ (em mM)			
Temperatura de anelamento		1,5	2	2,5	3
	56°C				
58°C			X	XX⊗	⊗
60°C		X		XX	

X = 33 *ng* cDNA
 ⊗ = 100 *ng* cDNA

Legenda: Acima: Concentrações de MgCl₂ (em mM) testadas. À esquerda: temperaturas (em graus Celsius) testadas.

Através da técnica de RT-PCR, obtivemos a amplificação dos transcritos constitutivos de todos os genes, e visualizamos também alguns dos possíveis transcritos alternativos. Os *amplicons* detectados foram seqüenciados a fim de confirmar a presença de transcritos constitutivos ou variantes. As amostras utilizadas no seqüenciamento foram obtidas de três maneiras diferentes: diretamente a partir do produto de RT-PCR, a partir de

reações de PCR efetuadas com o material genético presente nas bandas evidenciadas no gel de agarose, ou a partir dos vetores aonde o *amplicon* foi clonado.

Para os casos aonde diferentes bandas foram amplificadas e visualizadas, porém não estavam isoladas o suficiente para permitir uma boa reação de seqüenciamento, tentamos, então, realizar um novo PCR a partir de material genético presente nas bandas do gel de agarose. Quando este procedimento não produziu resultados eficazes, realizamos a separação de bandas a partir do gel de agarose. Recortamos as bandas do gel após a eletroforese e realizamos sua purificação, obtendo, por fim, somente o material genético presente em cada banda específica.

Nos casos em que as técnicas anteriormente descritas ainda foram incapazes de separar os diferentes *amplicons*, como ocorreu com os genes OXSM e C18orf19 que apresentaram *amplicons* de tamanhos próximos, utilizamos a técnica de clonagem desses fragmentos em vetor pT257 R/T e transformação por eletroporação na bactéria *E. coli* DH10B. Através desta técnica, o isolamento de diferentes fragmentos amplificados por RT-PCR ocorre a partir da utilização de quantidades de insertos e vetores (3:1) que, teoricamente, permitem que apenas um tipo de inserto esteja presente em cada colônia bacteriana. A partir disso, realizamos reações de PCR de colônia. Para confirmar a presença do inserto, realizamos a purificação com o uso da técnica de “*Miniprep*” partindo das colônias que deram indícios de conter os insertos desejados, obtendo, por fim, uma solução de DNA plasmidial. Realizamos a digestão desse produto com enzimas de restrição, que clivaram os plasmídeos liberando os insertos, o que nos permitiu identificar através da análise por eletroforese em gel de agarose quais colônias continham os fragmentos de interesse. Os produtos de digestão que continham os insertos desejados foram seqüenciados.

Consideramos que a utilização de tais técnicas possibilitariam a amplificação e visualização de todos os transcritos flanqueados por nossos iniciadores que estivessem

presentes no RNA total extraído das linhagens, ou seja, que obtivemos êxito em amplificar todos os transcritos constitutivos e variantes existentes nas células para os trechos dos genes de interesse.

Listaremos, a seguir, os resultados obtidos para cada gene estudado.

5.3. Resultados obtidos para cada gene estudado

Apresentam-se a seguir as representações e os resultados relativos a cada um dos sete genes estudados. A princípio, apresentamos uma tabela contendo as principais características descritas para cada gene, seguido da representação de todas as variantes já descritas para o gene, do trecho amplificado por PCR, e da localização e tamanho das *probesets* utilizadas pela técnica de “*Exon Array*” da Affymetrix®, dados provenientes do “*UCSC Genome Browser*” (<http://genome.ucsc.edu/>). As linhas representam os trechos intrônicos, e as barras os trechos exônicos. As *probesets* que apresentaram “*Splicing Index*” alterado, de acordo com nossa predefinição, encontram-se destacadas nos esquemas, e os índices podem ser visualizados nos quadros azuis. Abaixo da representação proveniente do “*UCSC Genome Browser*”, temos um esquema do que esperamos encontrar como transcrito constitutivo e como alternativo de acordo com os dados gerados. Ao lado de cada transcrito apresenta-se o tamanho do transcrito constitutivo e uma estimativa de tamanho do transcrito alternativo feita utilizando-se os tamanhos das *probesets* aonde houve variação do “*Splicing Index*”. As setas em preto e vermelho representam, respectivamente, o início e fim do trecho amplificado e do exon que apresentaria evento de *splicing* alternativo, e a barra em vermelho representa o exon aonde ocorre o evento. Por fim, apresentamos o gel de agarose no qual foi possível visualizar todos os transcritos amplificados para cada linhagem celular.

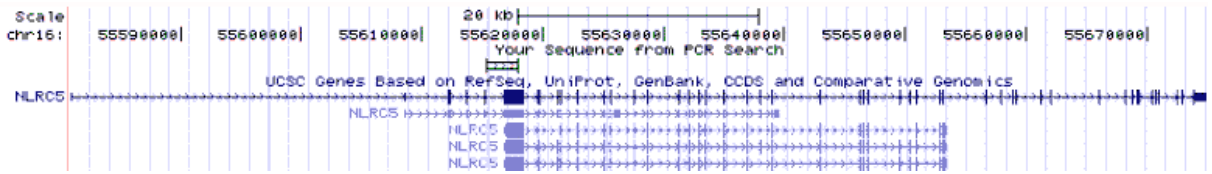
Genes que apresentaram alterações no *Exon Array* condizentes com eventos de *splicing* alternativo de exclusão parcial de exon

Os genes NLRC5 e KIF16B apresentaram, no trecho gênico amplificado, *probesets* alteradas no meio do exon, sendo que as *probesets* vizinhas tendiam a um valor não relevante. O valor positivo do índice para o gene NLRC5 indica que o trecho em questão provavelmente é mais incluso nas linhagens que não contêm a mutação que na linhagem mutada, incitando que a mutação pode estar envolvida na promoção de um evento de *splicing* alternativo que gere exclusão parcial do exon. Já para o gene KIF16B, os valores de “*Splicing Index*” foram negativos, sugerindo que o trecho da *probeset* está mais incluso na linhagem mutada que nas demais, e, deste modo, a mutação poderia estar envolvida em um menor acometimento do evento de *splicing* alternativo que exclui o trecho em questão. A partir da eletroforese dos produtos de RT-PCR, obtivemos os géis em questão, nos quais podemos observar a amplificação dos transcritos constitutivos, confirmados por reações de seqüenciamento. Nenhuma banda com tamanho próximo ao estimado para os transcritos alternativos foi visualizada através de nossas análises.

5.3.1. Gene NLRC5

GENE	NLRC5
Nomenclatura oficial	NLR family, CARD domain containing 5
Organismo	<i>Homo sapiens</i>
Também conhecido por	NOD4, NOD27, CLR16.1, FLJ21709 e FLJ39711
Funções moleculares descritas no Gene Ontology (GO)	Ligante de nucleotídeo, ligante de proteína, ligante de ATP
Localização celular proteica	Citoplasma

Figura 5.3.1.a. Representação do gene NLRC5 e do trecho amplificado por PCR.



Legenda: Em azul: representações dos diversos variantes já descritos para o gene. Em preto: trecho do gene a ser amplificado.

Figura 5.3.1.b. Detalhamento do trecho amplificado, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos para o gene NLRC5.

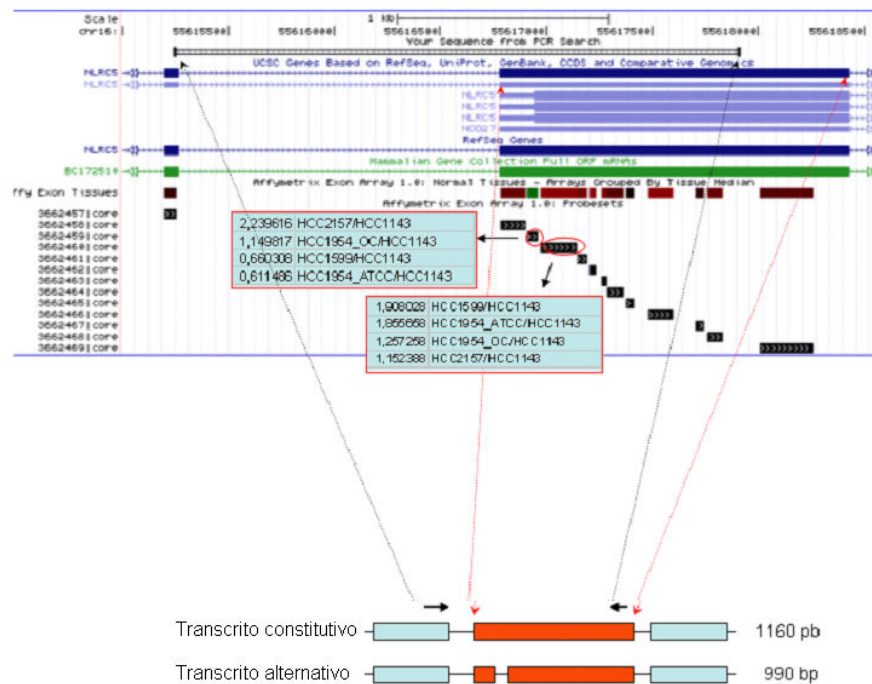
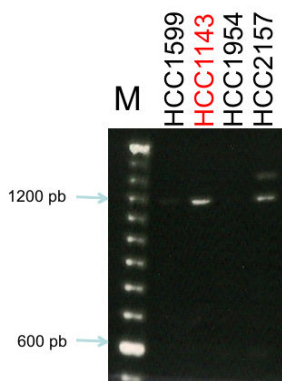


Figura 5.3.1.c. Gel de agarose da eletroforese dos produtos de RT-PCR do gene NLRC5.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados. M = marcador.

5.3.2. Gene KIF16B

GENE	KIF16B
Nomenclatura oficial	kinesin family member 16B
Organismo	<i>Homo sapiens</i>
Também conhecido por	SNX23, C20orf23 e KISC20ORF
Participação em processos biológicos descritas no Gene Ontology (GO)	Movimento por microtúbulos, comunicação celular
Funções moleculares descritas no Gene Ontology (GO)	Ligante de nucleotídeo, atividade motora, atividade motora de microtúbulos, ligante de proteínas, ligante de ATP
Localização celular proteica	Microtúbulo

Figura 5.3.2.a. Representação do gene KIF16B e do trecho amplificado por PCR.

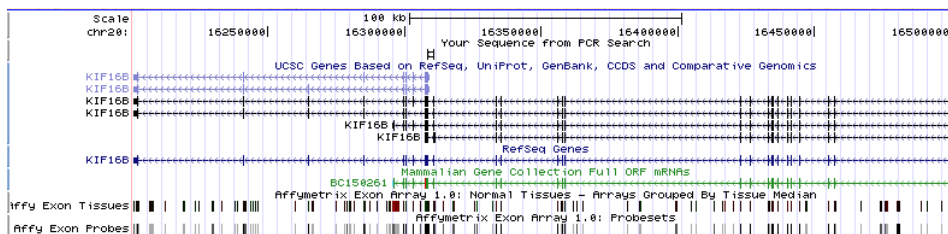


Figura 5.3.2.b. Detalhamento do trecho amplificado, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos para o gene KIF16B.

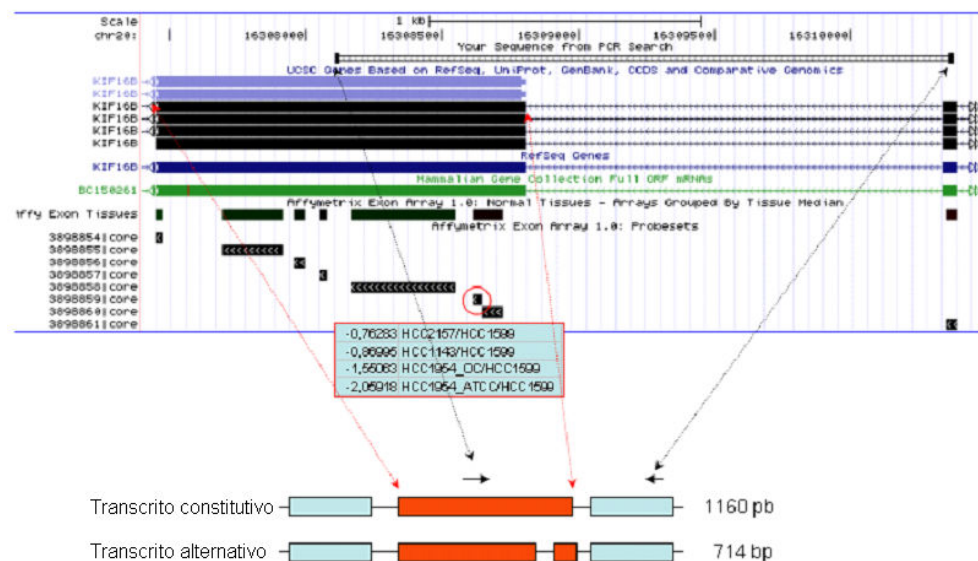
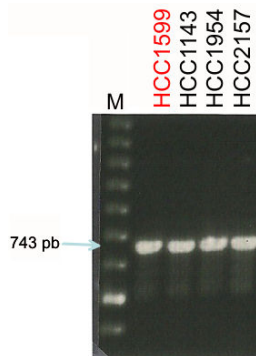


Figura 5.3.2.c. Gel de agarose da eletroforese dos produtos de RT-PCR do gene KIF16B.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados.

M = marcador.

Genes que apresentaram alterações no *Exon Array* condizentes com eventos de sítio alternativo de *splicing* 3' ou 5'

Os demais genes validados apresentaram alterações em índices “*Splicing Index*” em *probesets* localizadas próximas às bordas 3' (OXSM, KIAA1826, FEM1C e C18orf19) ou 5' (EHBP1) dos exons, sugerindo a ocorrência de eventos de sítio alternativo de *splicing*. Em todos esses casos, o “*Splicing Index*” alterado apresentava valores positivos, o que significa que o trecho gênico em questão está mais incluso nas linhagens aonde a mutação está ausente que na linhagem mutada. Assim sendo, podemos relacionar a ocorrência da mutação com uma maior ocorrência de eventos de *splicing* alternativo.

Nas linhagens KIAA1826, FEM1C e EHBP1 houve amplificação somente do transcrito constitutivo, como pode ser visualizado nas figuras dos géis de agarose.

Na linhagem C18orf19, além do transcrito constitutivo, houve amplificação de mais duas bandas, uma relacionada ao transcrito alternativo esperado, outra de tamanho intermediário aos transcritos constitutivo e alternativo. Porém, a amplificação do transcrito alternativo não foi restrita à linhagem mutada, podendo ser visualizada também nas demais linhagens, embora, aparentemente, em uma menor quantidade. Assim sendo, neste caso a mutação não pode ser relacionada como causa da ocorrência do evento de *splicing* alternativo, mas pode ser um fator adicional promotor de sua ocorrência.

Na linhagem OXSM foram amplificados e sequenciados em todas as linhagens o transcrito constitutivo e um transcrito alternativo, ainda não descrito na literatura, com cerca de 180 pares de bases. Através da análise da sequência proteica do novo variante por intermédio do banco de dados “Pfam” (<http://pfam.sanger.ac.uk/>), averiguamos que o novo variante apresenta uma ruptura em um domínio proteico da família “*Betha-ketoacyl synthase*”, cuja representação pode ser visualizada na figura 5.3.7.c.. Conseguimos visualizar também uma banda com tamanho que remete ao variante que esperávamos encontrar na linhagem mutada, porém, ainda não obtivemos sucesso no sequenciamento deste transcrito.

5.3.3. KIAA1826

GENE	KIAA1826
Nomenclatura oficial	KIAA1826
Organismo	<i>Homo sapiens</i>
Também conhecido por	
Participação em processos biológicos descritas no Gene Ontology (GO)	Nada descrito
Funções moleculares descritas no Gene Ontology (GO)	Nada descrito
Localização celular proteica	Núcleo, membrana

Figura 5.3.3.a. Representação do gene KIAA1826, do trecho amplificado por PCR, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos.

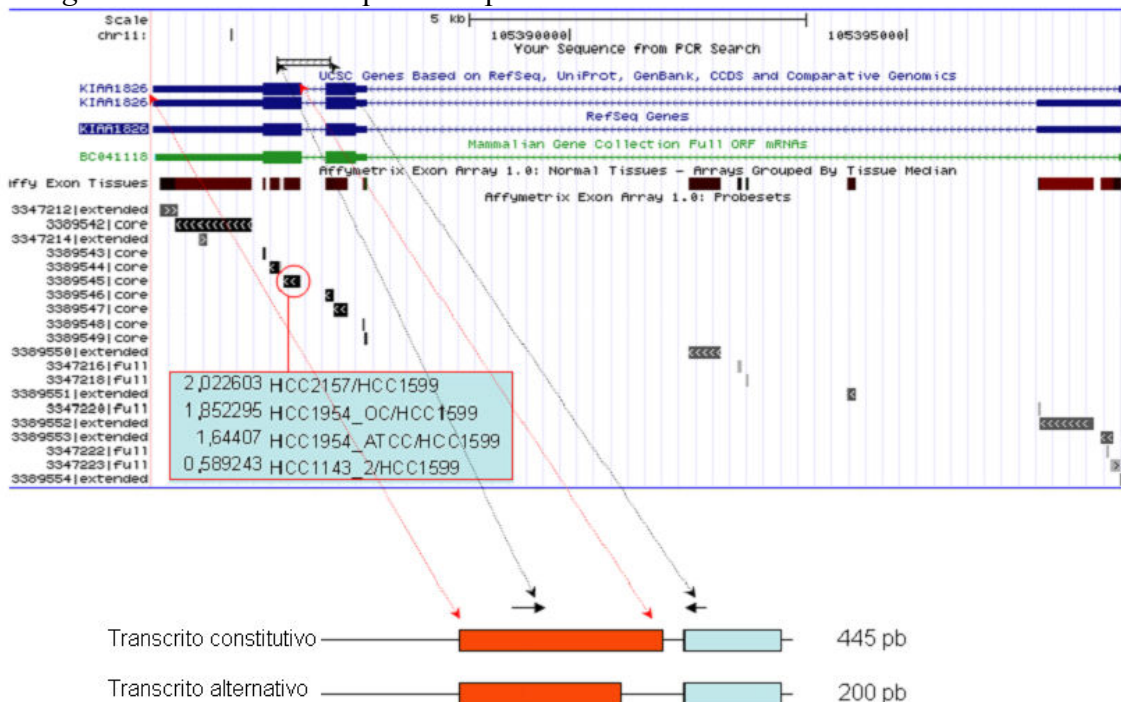
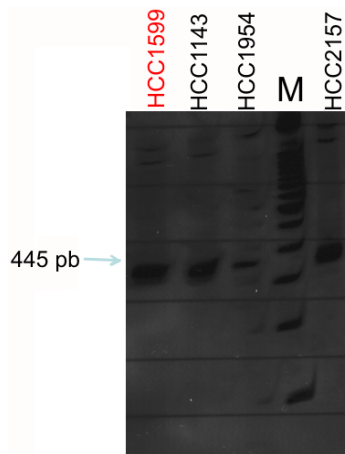


Figura 5.3.3.b. Gel de agarose da eletroforese dos produtos de RT-PCR do gene KIAA1826.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados. M = marcador.

5.3.4. FEM1C

GENE	FEM1C
Nomenclatura oficial	Fem-1 homolog c (<i>C. elegans</i>)
Organismo	<i>Homo sapiens</i>
Também conhecido por	FEM1A, KIAA1785 e EUROIMAGE686608
Participação em processos biológicos descritas no Gene Ontology (GO)	Nada descrito
Funções moleculares descritas no Gene Ontology (GO)	Nada descrito
Localização celular proteica	Nada descrito

Figura 5.3.4.a. Representação do gene FEM1C, do trecho amplificado por PCR, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos.

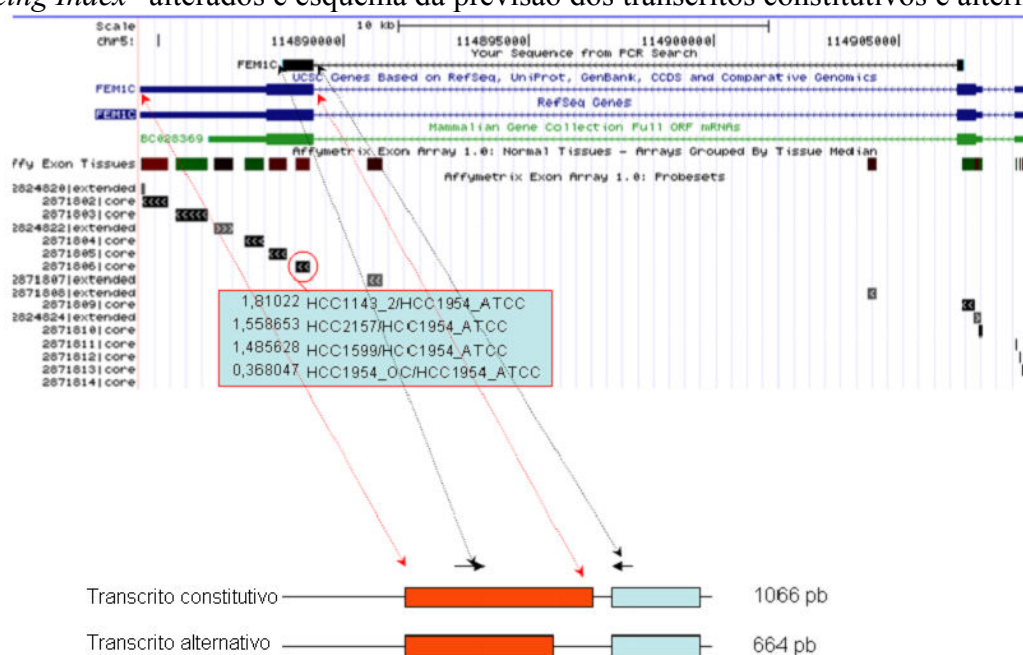
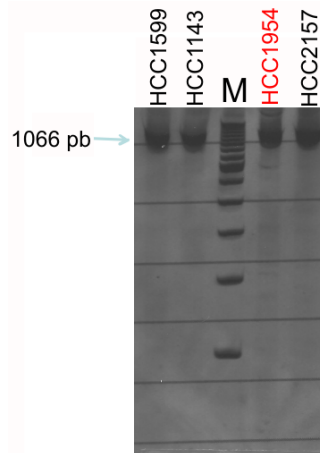


Figura 5.3.4.b. Gel de agarose da eletroforese da PCR a partir do gel de agarose da eletroforese do RT-PCR do gene FEM1C.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados. M = marcador.

5.3.5. EHBP1

GENE	EHBP1
Nomenclatura oficial	EH domain binding protein 1
Organismo	<i>Homo sapiens</i>
Também conhecido por	HPC12, NACSIN e KIAA0903
Participação em processos biológicos descritas no Gene Ontology (GO)	Nada descrito
Funções moleculares descritas no Gene Ontology (GO)	Nada descrito
Localização celular proteica	Nada descrito

Figura 5.3.5.a. Representação do gene EHBP1, do trecho amplificado por PCR, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos.

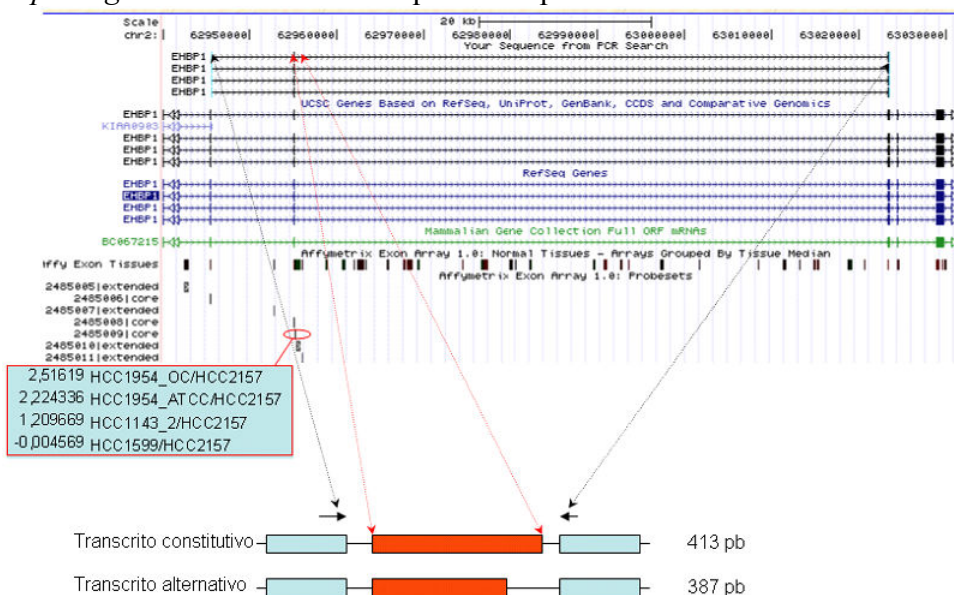
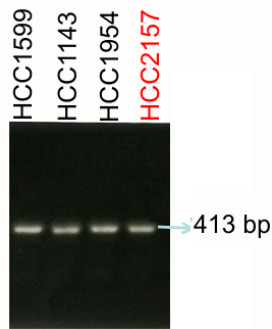


Figura 5.3.5.b. Gel de agarose da eletroforese do RT-PCR do gene EHBP1.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados.

5.3.6. C18orf19

GENE	C18ORF19
Nomenclatura oficial	chromosome 18 open reading frame 19
Organismo	<i>Homo sapiens</i>
Também conhecido por	HsT2329 e MGC24180
Participação em processos biológicos descritas no Gene Ontology (GO)	Nada descrito
Funções moleculares descritas no Gene Ontology (GO)	Nada descrito
Localização celular proteica	membrana

Figura 5.3.6.a Representação do gene C18orf19, do trecho amplificado por PCR, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos.

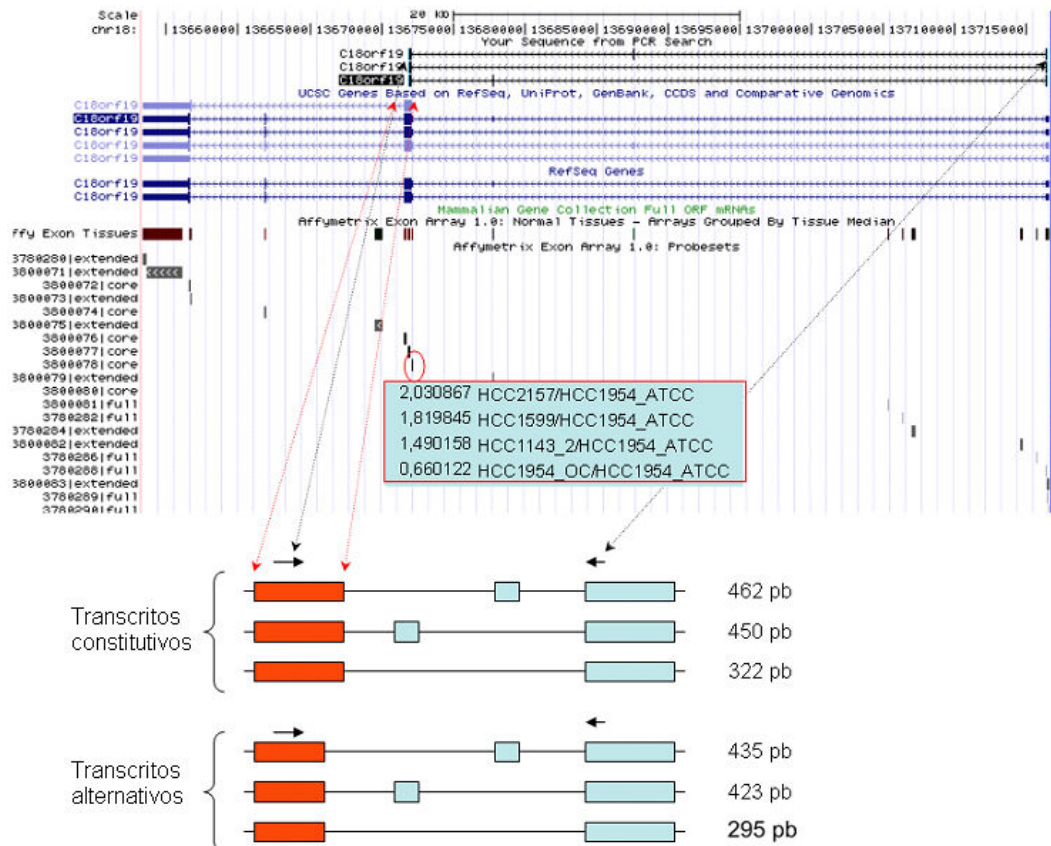
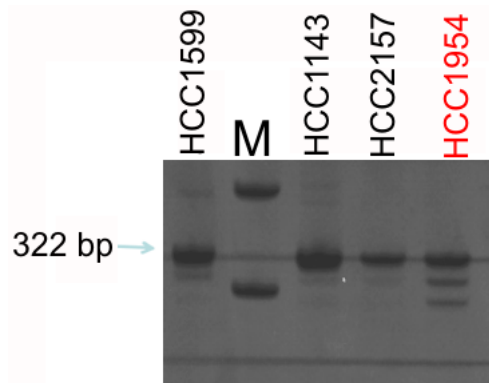


Figura 5.3.6.b. Gel de agarose da eletroforese do PCR de colônia do gene C18orf19.



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados. M = marcador.

5.3.7. OXSM

GENE	OXSM
Nomenclatura oficial	3-oxoacyl-ACP synthase, mitochondrial
Organismo	<i>Homo sapiens</i>
Também conhecido por	KS, FASN2D e FLJ20604
Participação em processos biológicos descritas no Gene Ontology (GO)	Processo de biossíntese de ácidos graxos, metabolização de acyl-CoA, processos de biossíntese de lipídeos
Funções moleculares descritas no Gene Ontology (GO)	Atividade catalítica, atividade de 3-oxoacyl-[acyl-carrier-protein] synthase, atividade de aciltransferase, atividade de transferase.
Localização celular proteica	Mitocôndria

Figura 5.3.7.a. Representação do gene OXSM, do trecho amplificado por PCR, índices “Splicing Index” alterados e esquema da previsão dos transcritos constitutivos e alternativos.

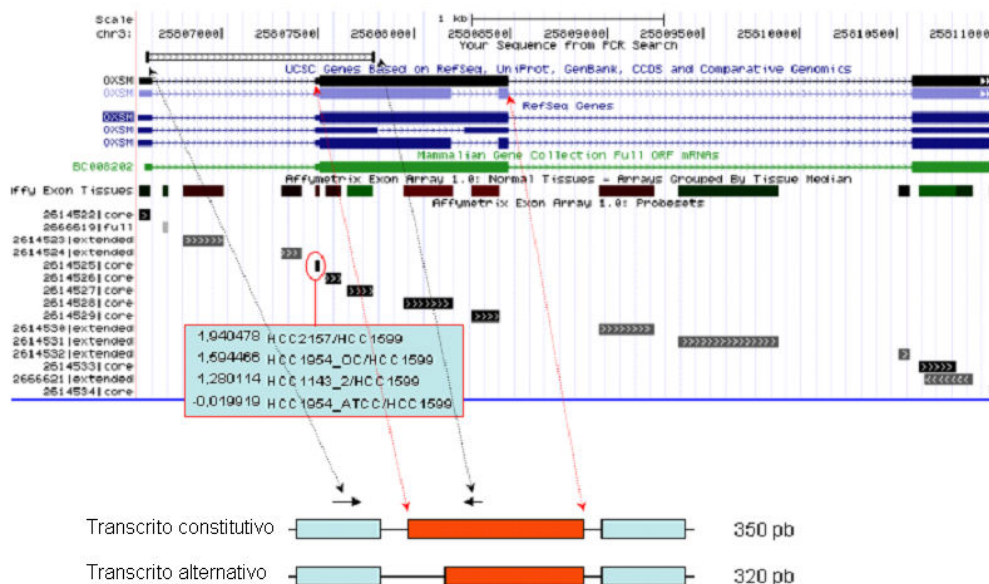
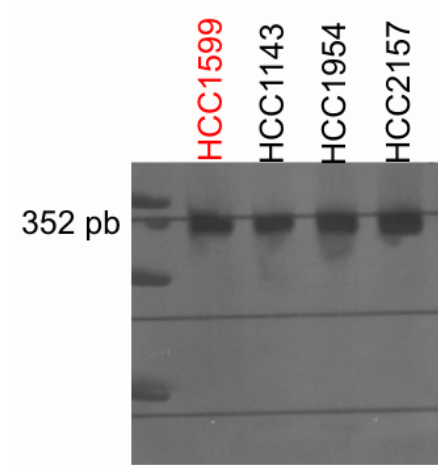
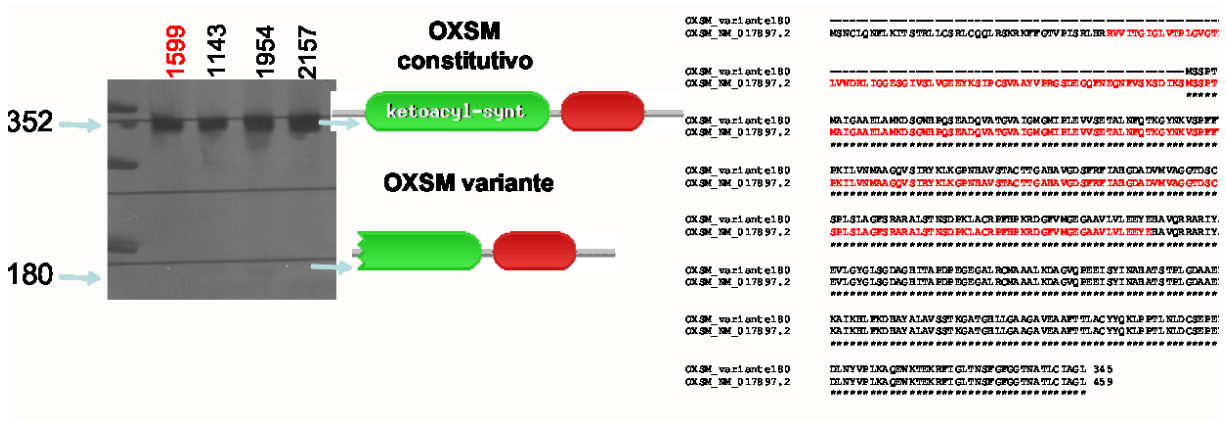


Figura 5.3.3.b. Gel de agarose da eletroforese da PCR do gene OXSM..



Legenda: Acima: Linhagens celulares, em vermelho a linhagem aonde ocorre a mutação. Ao lado, os tamanhos de transcritos (em pares de bases) encontrados.

Figura 5.3.3.c. Análise comparativa das sequências gênicas e proteicas dos transcritos constitutivo e alternativo do gene OXSM através do banco de dados “Pfam”.



Legenda: À esquerda, o gel de agarose com as bandas dos transcritos amplificados. No meio, a representação criada pelo “Pfam” de como estariam os domínios proteicos conhecidos nos transcritos constitutivo e alternativo. À direita, a complementaridade entre as sequências proteicas dos transcritos constitutivo e variante.

6. CONCLUSÕES

“Although the DNA sequence of the human genome is known, it will probably take at least a decade for human to identify every gene and determine the precise amino acid sequence of the protein it produces. Yet the cells in our body do this thousands of times a second.”
(Bruce Alberts)

Neste trabalho, investigamos, em linhagens tumorais de mama, se mutações de um par de bases localizadas próximas à bordas de *splicing* têm potencial de gerar ou alterar eventos de *splicing* alternativo. Foram selecionados sete genes para validação experimental, e, destes, apenas dois genes apresentaram variantes de *splicing* no trecho amplificado: o gene C18orf19, que apresentou dois transcritos variantes, que foram amplificados em todas as linhagens; e o gene OXSM, que apresentou um transcrito variante presente em todas as linhagens e um variante aparentemente específico à linhagem mutada. Assim sendo, somente em 14% (gene OXSM) dos casos selecionados para validação haveria possibilidade de geração de um evento de *splicing* alternativo a partir de uma mutação pontual, o que consideramos um baixo percentual de comprovação experimental. No caso específico do gene C18orf19, a diferença de intensidade entre as bandas amplificadas comparativamente entre a linhagem com a mutação e as demais linhagens, sugere a hipótese de que a mutação esteja influenciando um evento de *splicing* alternativo, promovendo-o mais que o ocorrido em linhagens sem mutação, porém, são necessário experimentos mais detalhados para comprovar tal situação. Em suma, concluímos que, dentro de nossa seleção de casos, as mutações pontuais têm uma baixa influência na geração e promoção de eventos de *splicing* alternativo.

Adicionalmente, encontramos 4 transcritos variantes não descritos na literatura dentre 7 trechos gênicos estudados - sendo ainda que, um destes, em análises preliminares, apresentou um importante domínio proteico trincado em sua estrutura. Sob tal perspectiva, podemos prever que ainda há uma subestimação do número e da importância celular de variantes de *splicing*, reafirmando, então, a premissa que o estudo de eventos de *splicing*

alternativo ainda é uma vasta área a ser explorada, estando altamente propícia a novas descobertas e provavelmente tendo uma alta importância nas alterações do metabolismo celular.

APÊNDICE

A.1. Gel de agarose 1%

A.2. Reações de PCR para genes controle

A.2.1. hMLH1

A.2.2. Notch2

A.3. Reações de PCR para genes escolhidos

A.3.1. NLRC5

A.3.2. KIF16B

A.3.3. OXSM

A.3.4. KIAA1826

A.3.5. FEM1C

A.3.6. C18ORF19

A.3.7. EHBP1

A.1. Gel de agarose 1%

O gel de agarose foi preparado para 25mL a cada placa de eletroforese, e teve a composição de 1% agarose em 1x TAE (Tris-Acetato-EDTA. Os ácidos nucleicos foram visualizados usando o “GelRedTM” (Biotium, Inc.), um tipo de ácido nucleico fluorescente, e visualizados por intermédio de exposição à luz ultra-violeta.

A.2. Reações de PCR para os genes controle

A.2.1. Gene mutL homolog 1 (hMLH1)

O RNA tratado com DNase de cada linhagem foi amplificado utilizando os iniciadores específicos para o gene, detalhados em Materiais e Métodos. Como controle positivo foi utilizado DNA genômico. Os produtos amplificados foram visualizados em gel de 8% poliacrilamida corado com nitrato de prata.

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	0,75 µL
dNTP's (1,25mM)	01,0 µL
Primer Forward (10µM)	0,8 µL
Primer Reverse (10µM)	0,8 µL
Taq Platinum	0,2 µL
H ₂ O Miliq	18 µL
Amostra RNA	1,0 µL
Total	25,0 µL

Ciclagem:
 95°C – 5min
 95°C – 45s
 60°C – 45s
 72°C – 1min
 72°C – 7min
 4°C – final da reação

} 35 ciclos

A.2.2. Notch2 (Notch homolog 2)

O cDNA foi amplificado com os iniciadores específicos detalhados anteriormente na seção Materiais e Métodos. Como controle positivo foi utilizado cDNA da linhagem HB4a (tecido de mama normal) previamente sintetizado e testado. Os produtos da amplificação foram analisados em gel de 8% poliacrilamida corado com prata.

Reagentes	Quantidade
Buffer 10x	2,0 µL
MgCl ₂ (50mM)	0,6 µL
dNTP's (1,25mM)	0,5 µL
Primer Forward (10µM)	0,8 µL
Primer Reverse (10µM)	0,8 µL
Taq Platinum	0,2 µL
H ₂ O Miliq	11,4 µL
Amostra cDNA	1,0 µL
Total	20,0 µL

Ciclagem:
 95°C – 5 min
 95°C – 45 s
 60°C – 45 s
 72°C – 1 min
 72°C – 7 min
 4°C – final da reação

} 35 ciclos

A.3. Reações de PCR para os genes de interesse

O cDNA de cada linhagem foi amplificado utilizando os iniciadores específicos para cada gene, detalhados em “Materiais e métodos”. As amplificações consideradas ótimas de cada gene de interesse requisitaram condições químicas e físicas que diferiram individualmente. As reações químicas e as condições de ciclagem em equipamento de PCR para cada gene encontram-se demonstradas a seguir.

A.3.1. NLRC5

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,0 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,9 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 min
 60°C – 30 s
 72°C – 1 min 10 s
 72°C – 6 min
 4°C – final da reação

} 35 ciclos

A.3.2. KIF16B

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,0 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,9 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 min
 60°C – 30 s
 72°C – 1 min 10 s
 72°C – 6 min
 4°C – final da reação

} 35 ciclos

A.3.3. OXSM

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	0,75 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	19,15 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 s
 58°C – 30 s
 72°C – 1 min 10 s
 72°C – 6 min
 4°C – final da reação

} 35 ciclos

A.3.4. KIAA1826

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,25 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,65 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 s
 60°C – 30 s
 72°C – 1 min 10 s } 35 ciclos
 72°C – 6 min
 4°C – final da reação

A.3.5. FEM1C

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,5 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,4 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 s
 60°C – 30 s
 72°C – 1 min 10 s } 35 ciclos
 72°C – 6 min
 4°C – final da reação

A.3.6. C18ORF19

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,5 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,4 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 s
 60°C – 30 s
 72°C – 1 min 10 s } 35 ciclos
 72°C – 6 min
 4°C – final da reação

A.3.7. EHBP1

Reagentes	Quantidade
Buffer 10x	2,5 µL
MgCl ₂ (50mM)	1,0 µL
dNTP's (10mM)	0,5 µL
Primer Forward (10mM)	0,5 µL
Primer Reverse (10mM)	0,5 µL
Taq Platinum	0,1 µL
H ₂ O Miliq	18,9 µL
Amostra cDNA	1,0 µL
Total	25,0 µL

Ciclagem:

94°C – 2 min
 94°C – 30 s
 60°C – 30 s
 72°C – 1 min 10 s } 35 ciclos
 72°C – 6 min
 4°C – final da reação

BIBLIOGRAFIA

1. World Health Organization. Fact sheet N°297. February 2009. 2009. Ref Type: Internet Communication
2. Brasil, M. d. S., Secretaria de Atenção à Saúde, Instituto Nacional de Câncer, & Coordenação de Prevenção e Vigilância de Câncer. (2007).
3. Danaei, G., Vander, H. S., Lopez, A. D., Murray, C. J., & Ezzati, M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet* **366**, 1784-1793 (2005).
4. Coleman, W. B. & Tsongalis, G. J. Molecular mechanisms of human carcinogenesis. *EXS* 321-349 (2006).
5. Pearson, P. L. & van der Luijt, R. B. The genetic analysis of cancer. *J.Intern.Med.* **243**, 413-417 (1998).
6. Yeo, G., Hoon, S., Venkatesh, B., & Burge, C. B. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc.Natl.Acad.Sci.U.S.A* **101**, 15700-15705 (2004).
7. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113 (2007).
8. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274 (2006).
9. Futcher, B. & Leatherwood, J. K. Molecular biology: Bound to splice. *Nature* **455**, 885-886 (2008).
10. Tempfer, C. B., Hefler, L. A., Schneeberger, C., & Huber, J. C. How valid is single nucleotide polymorphism (SNP) diagnosis for the individual risk assessment of breast cancer? *Gynecol.Endocrinol.* **22**, 155-159 (2006).
11. Bhangale, T. R., Rieder, M. J., Livingston, R. J., & Nickerson, D. A. Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. *Hum.Mol.Genet.* **14**, 59-69 (2005).
12. Weber, J. L. *et al.* Human diallelic insertion/deletion polymorphisms. *Am.J.Hum.Genet.* **71**, 854-862 (2002).
13. Mills, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **16**, 1182-1190 (2006).
14. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-876 (2008).
15. Costantini, M. & Bernardi, G. Mapping insertions, deletions and SNPs on Venter's chromosomes. *PLoS.One.* **4**, e5972 (2009).
16. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat.Genet.* **39**, 1217-1224 (2007).
17. Kim, J. I. *et al.* A highly annotated whole-genome sequence of a Korean individual. *Nature* (2009).
18. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat.Methods* **6**, 377-382 (2009).
19. Kato, K. Impact of the next generation DNA sequencers. *Int.J.Clin.Exp.Med.* **2**, 193-202 (2009).
20. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933 (2001).
21. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158 (2007).
22. STRACHAN, T. & READ, A. P. *Genética Molecular Humana* Art Med, Porto Alegre (2002).
23. LODISH, H. & *et al.* *Molecular Cell Biology* WH Freeman,(2003).
24. ALBERTS, B. & *et al.* *Molecular Biology of the cell.* Garland Science,(2002).
25. Nilsen, T. W. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25**, 1147-1149 (2003).
26. House, A. E. & Lynch, K. W. Regulation of alternative splicing: more than just the ABCs. *J.Biol.Chem.* **283**, 1217-1221 (2008).
27. PIERCE, B. A. *Genetics – A conceptual Approach.* WH Freeman,(2008).
28. LEWIN, B. *Genes*, **VII**, Oxford University Press,(2000).
29. Venables, J. P. *et al.* Cancer-associated regulation of alternative splicing. *Nat.Struct.Mol.Biol.* (2009).
30. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA.* **14**, 802-813 (2008).

31. Wachtel, C. & Manley, J. L. Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol.Biosyst.* **5**, 311-316 (2009).
32. Ladd, A. N. & Cooper, T. A. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.* **3**, reviews0008 (2002).
33. Venables, J. P. Aberrant and alternative splicing in cancer. *Cancer Res.* **64**, 7647-7654 (2004).
34. Schaal, T. D. & Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol.Cell Biol.* **19**, 1705-1719 (1999).
35. Sammeth, M., Foissac, S., & Guigo, R. A general definition and nomenclature for alternative splicing events. *PLoS.Comput.Biol.* **4**, e1000147 (2008).
36. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141-2144 (2003).
37. Xu, Q., Modrek, B., & Lee, C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754-3766 (2002).
38. Matlin, A. J., Clark, F., & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nat.Rev.Mol.Cell Biol.* **6**, 386-398 (2005).
39. Yeo, G., Holste, D., Kreiman, G., & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biol.* **5**, R74 (2004).
40. Ben Dov, C., Hartmann, B., Lundgren, J., & Valcarcel, J. Genome-wide analysis of alternative pre-mRNA splicing. *J.Biol.Chem.* **283**, 1229-1233 (2008).
41. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476 (2008).
42. Kornblihtt, A. R., de la, M. M., Fededa, J. P., Munoz, M. J., & Nogues, G. Multiple links between transcription and splicing. *RNA.* **10**, 1489-1498 (2004).
43. Kurmangaliyev, Y. Z. & Gelfand, M. S. Computational analysis of splicing errors and mutations in human transcripts. *BMC.Genomics* **9**, 13 (2008).
44. Kralovicova, J. & Vorechovsky, I. Global control of aberrant splice-site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. *Nucleic Acids Res.* **35**, 6399-6413 (2007).
45. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat.Rev.Genet.* **8**, 749-761 (2007).
46. Kim, E., Goren, A., & Ast, G. Alternative splicing and disease. *RNA.Biol.* **5**, 17-19 (2008).
47. Srebrow, A. & Kornblihtt, A. R. The connection between splicing and cancer. *J.Cell Sci.* **119**, 2635-2641 (2006).
48. Kirschbaum-Slager, N., Lopes, G. M., Galante, P. A., Riggins, G. J., & de Souza, S. J. Splicing factors are differentially expressed in tumors. *Genet.Mol.Res.* **3**, 512-520 (2004).
49. Ghigna, C., Valacca, C., & Biamonti, G. Alternative splicing and tumor progression. *Curr.Genomics* **9**, 556-570 (2008).
50. Song, S. *et al.* Comparative component analysis of exons with different splicing frequencies. *PLoS.One.* **4**, e5387 (2009).
51. GeneChip® Exon Array Design. *Identifying and Validating Alternative Splicing Events*(2006).